

Table 7. Recommended process for documentation of performance parameters and comparability of two different bioassessment methods

[Five reference sites are assumed in this layout, but one could have a minimum of three sites for each region]

	Region 1				Region 2			
	Reference numbers 1–5		Impaired or test site		Reference numbers 1–5		Impaired or test site	
	Method 1, mean variance	Method 2, mean variance	Method 1	Method 2	Method 1, mean variance	Method 2, mean variance	Method 1	Method 2
Metric ₁	$\mu_1 \pm S_1$	$\mu_2 \pm S_2$	m	p	$a_1 \pm d_1$	$a_2 \pm d_2$	c	q
.								
.								
Metric _n								
Assessment score.	$\chi_1 \pm q_1$	$\chi_2 \pm q_2$	z	v	$b_1 \pm f_1$	$b_1 \pm f_2$	e	r

The following comparisons refer to the parameters specified above and are designed to yield various performance characteristics of a biological-field-collection method.

- Compare s_1 with s_2 for a given metric to determine relative precision of the metric for the two methods and an unimpaired condition.
- Compare s_1 with d_1 and s_2 with d_2 to determine how metric variability may change with a region. A relatively high variability in a given metric within a region or compared with another region for the same method would suggest a certain performance range and bias for the metric.
- Compare $m\mu_1$ with $p\mu_2$ to determine discriminatory power of a given metric by using the two methods in region 1. A ratio closer to 1.0 would signify little difference in the metric between an impaired site and the reference condition in region 1 for that method. The utility of the metric would be questionable in this case. Do the same type of analysis by comparing c/a_1 and q/a_2 for region 2.
- Compare $m\mu_1$ with c/a_1 and $p\mu_2$ with q/a_2 to determine relative discriminatory power, performance range, and bases of a given metric and sampling method across regions. A similar ratio across regions for a given metric may indicate the robustness of the method and the metric. A ratio near 1.0 in one region and not in another for a given method and metric would indicate possible utility limitations or a limited performance range for that metric.
- Compare q_1 with q_2 and f_1 with f_2 to determine overall method variability at unimpaired sites in each region. High variability in the score for one method compared to another method in a given region would suggest lack of comparability and (or) different applicable data-quality operations for the two methods.
- Compare q_1 with f_1 and q_2 with f_2 to determine relative variability in assessment scores in the two regions. A consistently low score variability for a given method across regional reference sites would suggest method rigor and potential sensitivity.
- Compare resultant scores for a given method and region deleting apparently variable or insensitive metrics to determine metric redundancy and to determine relative discriminatory power at impaired sites.
- Individual assessment scores for reference sites and impaired sites within each region can be compared between methods by using regression to determine if there is a systematic relation in scores between the two methods.